

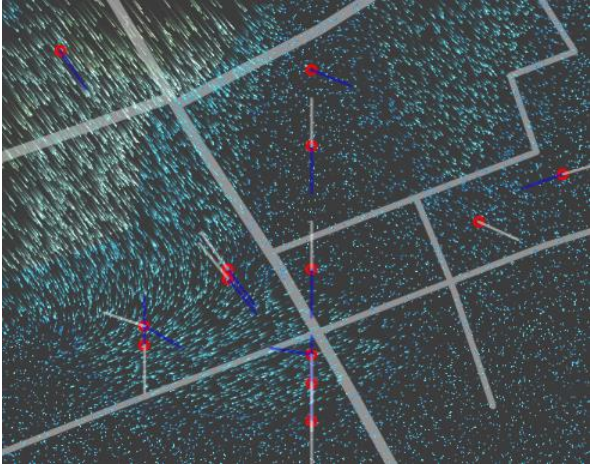
本周工作：

VAUD	9 月 10 号前提交（计划本月）	继续找小错误
语意轨迹	9 月 15 号	修改讨论，设计视觉编码，实现
阿里项目		实现，讨论，总结问题，分三步解决
硕士论文	8 月底	撰写中
课程 PPT 例子添加	本月	完成
专利	走之前	改了一半了
大屏脚本	8 月底	已完成英文版

详情：

1. 阿里项目进度：

- 新的研究方向：讨论进度和方向，可以考虑进行轨迹聚类，然后把轨迹扩散出去的方向来生成轨迹流场。这样生成的流畅可能更能代表大部分轨迹的流动方向。这是一个小方向和小的优化。
- 进度方面，关会华进行了程序整改和绘制，整个效果还是可以的也让大师姐看了一下。存在的主要问题和解决方法也已经讨论得出



- 边界问题，流场有时会有很强的边界性，原因是距离衰减函数和多边形边界的冲突，向量并没有衰减完就不扩散了。下周我将调整这部分代码，自适应的扩散到应有的地方。
- 聚类问题，现在的聚类算法 `kmeans` 每次结果不同，需要考虑一种稳定的聚类算法，下周关会华会修改这部分代码
- 数据问题，这周对全城的数据进行了测试，10 分钟的数据。下周将会在上面两步完成后，通过马哥之前的文章对正确性进行验证。

2. 语义轨迹项目：

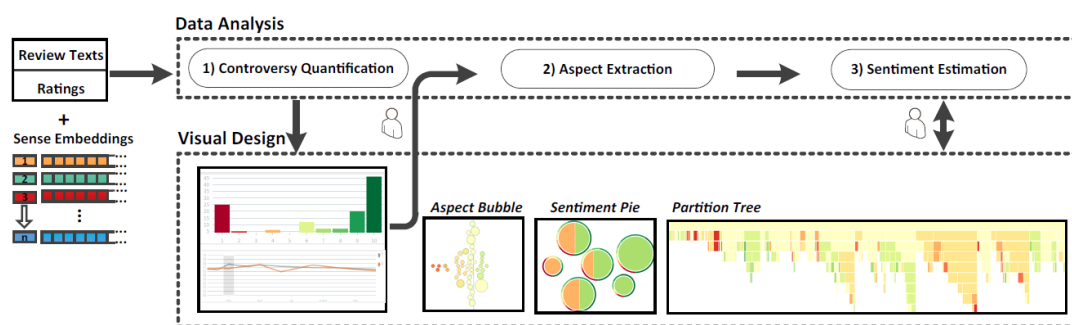
针对 **review** 制定了一套修改方向和计划，程序方面高盛杰开始进行修改。下周林柄如开始进行矩阵图的绘制。

之后可以根据总结计划写 **cover letter**。大改动的地方主要是动机，问题描述和 **case** 及评价方法。

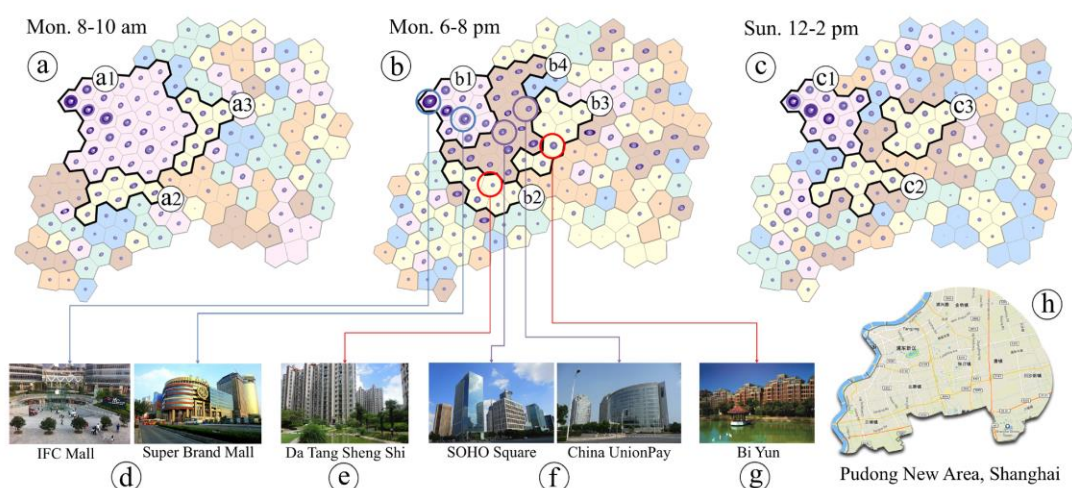
3. 大屏脚本，通过前两次介绍总结，我写了英文版本的 5 分钟和 10 分钟的大屏介绍方法，五分钟部分主要强调问题，动机，解决方法，没有实际操作方法，案例边讲解决方法边操作，但不介绍。10 分钟版本增加案例到找到出租车，但是也同样不介绍操作方法，只讲案例内容。
4. 课程 PPT 修改，修改了课程 PPT 交互章节中的例子，增加案例，修改图片和部分备注
5. 专利修改了前半部分（前 5 页），下周修改后半部分。
6. 侯总的开题报告撰写了后几节计划流程，并与嘉诚讨论了应用场景和写作方法。
7. 修改模板格式创建投送到 SI: VA 的论文
8. 澳洲回国总结整理，办理回国手续和报销整理，发送邮件联系出国办，下周去报销。美国出国手续校内流程完成。
9. 论文阅读：

### Exploring Controversy via Sentiment Divergences of Aspects in Reviews

本文通过层次聚类的方法，将文本中的热门话题进行层次划分，然后通过不同的表现形式对话题在时序上的变化进行了可视化。运用的数据是点评数据，整体的界面是一个集成性的系统，通过颜色编码评价的正负。文本方面的视觉编码可以参考本文



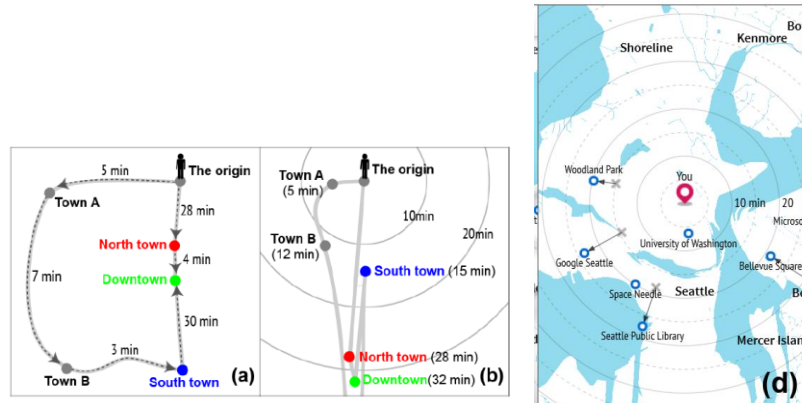
### MobiSeg: Interactive Region Segmentation Using Heterogeneous Mobility Data



本文将轨迹数据进行可视化编码，整合到六边形的区域内，应用非负矩阵分解算法捕捉潜在的行为模式，理解（其实是分类）人类移动行为，看这篇文章主要是想要看看有什么在六边形内编码的方法。他的可视化编码了不同数据的流入流出的量级，并且能够通过聚类，颜色来对数据进行探索。可以引用一下。

## Designing Interactive Distance Cartograms to Support Urban Travelers

本文对比了等距离地图和等轨迹时间地图的优劣，等轨迹时间地图，可能会产生过多的扭曲，不便于开车旅行，但是能对旅行时间成本一目了然。并对等轨迹时间地图，设计了两种交互方法，地理上下文的锚定投影技术，显示两种地图的结合，路网结构的简化展示，帮助用户进行导航。



### 附录：大屏 5 分钟版本

（语速稍快）

This is an integrated analysis platform for urban data. We collect urban data from many different application domains in a real middle size city with 7millions population. The amount of the data per day is about 300GB, and the entire timeline of the data we process is about 30 days in one month. The data itself is massive and heterogeneous, posing a huge challenge on data analysis and management.

To fuse and investigate the large amount of data, assist the security or government apartment to supervise and control the situation of the city. Our analysis platform support data querying, data analyzing, and data reasoning.

To provide data overview, data querying and data analyzing, the entire interface can be separated into two parts. The left part is an information panel, listing the data and their attributes we get. Including. The data sources are comprehensive covering human behaviors, the social, and static things like building. (这时候点几个统计图)

To analyze the data, the second part uses comprehensive views of displaying our daily life and support data querying and data reasoning. It is quite useful to do some management, plaining and decision making. When addressing a complex cross domain query, we can separate it into a series of simple query and information extraction across single data source. For example, when we want to find which place do people most likely lost their things. We can first query twitters which has the keyword 'lost'. And extract the geo-location information of the twitter to find what POI is near the twitter. (这时候开始选择微博，输入“丢失”，查询，显示，点选几个，点击街景，抽取位置，删除之前的条件，选择 POI，查询，显示，放大点击)

Each query operation can be divide into three stages, select data source, input query condition and define the Boolean condition of the conditions. When we get the query results, a situation awareness interface based on a map can help to show the detail of the results. By interact with the map, we can investigate the data and extract the information which we are interest to be a query condition to do the next step query.

In addition to the main views. We can integrate other technology to help manage and provide automatic analysis

of urban data. The left views using deep learning algorithm and KNN technology to analyze the functionality of the cell phone towers in the city. (这时候在地图上用朱博项目框选范围。)

## 附录：论文总结计划

1. I think the main problem is that the paper is written as a technique paper, but actually is an application paper... 读起来像是技术文章，需要描述问题，解决思路。
2. the book, called "Visual Analysis of Movement" is just one of the critical material. 书籍参考文献
3. The requirements stated in the introduction and motivate the paper is a repetition of the same idea, just phrased differently: "Semantic query over trajectory data is a type of fuzzy searching where trajectories related to the query keywords" and "Handling Uncertainty: The system should perform effective semantic query over uncertain data of the mobile phone recorded trajectories." Both are the same in my opinion, and they are both very high level. 文章的动机重复，层次太高。一个是通过模糊查询查询有关的轨迹，一个是通过语义查询不精确的轨迹。没说清楚
4. Chapter 4 and 5 review more the state-of-the-art in movement analysis and how it has been used in the current context. The figure 6 is a great demonstration of the visual encoding and mapping. However, there is a lack of evidence of a contribution. 我们需要从贡献点的角度出发。多给出我们的贡献，特点  
The paper presents novel ideas but lacks in giving motivation, justifying design decisions, and evaluating results. (R3)
5. The case-studies aim at demonstrating the use of the tool for a well-defined use case. I struggle with this a little bit:
  - if the questions and queries the user states, are very exact, how do you define the fuzzy-ness here? 查询条件选择，比较具体，可以考虑更加模糊
  - I am not quite sure about some of the discoveries: if there is a shopping center around a train station, or not, can not be a finding of the system, but rather a fact, that should have been known by any domain expert as well as any person in the area. So what is discovered here? 重点放在来商店的人大都来自那里，什么时间来
6. the paper read to me more as an application or system paper, than a technique paper. When rewriting in this format and adding appropriate evaluation, the paper might draw a lot of interest. 需要增加用户评测和适当的问题思路  
Besides the case studies, no other support for the effectiveness is reported (e.g., user study, expert feedback). (R1)
7. In the paper, I would clearly state what is the state-of-the-art and the contribution beyond it, whatever is not the contribution of the paper, should be emphasized this way. There should be also a clear statement of the actual contribution. 加强自己的可视化贡献的介绍
8. evaluation is by far not sufficient to demonstrate the usefulness of the approach. Current results of the case studies are rather trivial and do not require a sophisticated analysis. 用户评价，例子介绍，缺少分析  
I strongly advice to carry out a user study to evaluate if the design decisions are reasonable and the resulting approaches are useful and understandable/usable. (R3)
1. Results are too weak and not mapped to the claimed contributions of the paper. The contributions are not declared properly, do not represent a novel addition, and are not motivated by the reviewed related work. 没有强调贡献点

1. The presented solution is novel, but I have concerns regarding its **usefulness and soundness of the visual representation**. The evaluation of the approach just consists of a two small case studies that do not provide sufficient evidence on the effectiveness of the approach. 现有的视觉编码需要进行更改，更改思路从条件的选择，条件向量的自定义，表达方式的视觉编码，抽象图形的优化，视觉上的直观性这几点出发。需要通过案例说明有效性。需要充足的时间来调整风格和视觉效果
2. However, I was surprised to not find a discussion of how previous work reconstructed and visualized trajectories from mobile phone usage data. There exists relevant work that should be referenced and discussed, for instance: **之前工作的对比，介绍，和缺少的引用**
  - Sagl, Günther, Martin Loidl, and Euro Beinat. "A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic." *ISPRS International Journal of Geo-Information* 1, no. 3 (2012): 256-271.
  - Iovan, Corina, Ana-Maria Olteanu-Raimond, Thomas Couronné, and Zbigniew Smoreda. "Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies." In *Geographic Information Science at the Heart of Europe*, pp. 247-265. Springer International Publishing, 2013.
  - Smoreda, Zbigniew, Ana-Maria Olteanu-Raimond, and Thomas Couronné. "Spatiotemporal data from mobile phones for personal mobility assessment." In *Transport survey methods: best practice for decision making*, pp. 745-768. Emerald Group Publishing Limited, 2013.
  - Sagl, Guenther, Eric Delmelle, and Elizabeth Delmelle. "Mapping collective human activity in an urban environment based on mobile phone data." *Cartography and Geographic Information Science* 41, no. 3 (2014): 272-285.

by knowing the exact stop points of trips, one does not automatically know the following activity 关于轨迹的不确定性的研究。

  - C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis, and Z. Yan.: Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)*, 45 (4):42:1 - 42:32, 2013.
  - R. Krüger, D. Thom, and T. Ertl.: Visual analysis of movement behavior using web data for context enrichment. In *IEEE Pacific Visualization Symposium (PacificVis)*, pages 193 - 200. IEEE Computer Society, 2014a
  - Wei Zeng, ChiWing Fu, Stefan Arisona, Kwan Liu Ma: Visualizing the Relationship Between Human Mobility and Points of Interest. In *IEEE Transactions on Intelligent Transportation Systems* PP(99):114, 2017.

Furthermore as the paper deals with uncertainty and geovisualization, the authors could discuss how their thoughts align with MacEachren's taxonomy of uncertainty.

- MacEachren, Alan M., et al. "Visualizing geospatial information uncertainty: What we know and what we need to know." *Cartography and Geographic Information Science* 32.3 (2005): 139-160.

Another missing part is visual querying. As a main part of the paper concerns the finding of different situations and a query language for semantic trajectory data, the authors should not only discuss index structures but also approaches about textual/visual querying and trajectory querying. For example, I found

- F. Haag, R. Krüger, and T. Ertl.: VESpa: A pattern-based visual query language for event sequences. In *11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (IVAPP)*, volume 7, pages 48 - 59, 2016

It presents a visual query language for semantically enriched trajectory data, which looks very similar to the result patterns presentation in this work. The authors should clearly discuss similarities and differences.

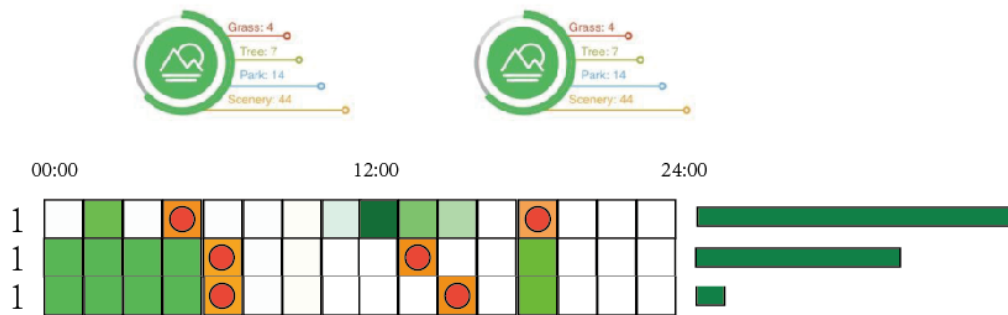
- MacEachren, Alan M., et al. "Visualizing geospatial information uncertainty: What we know and what we need to know." Cartography and Geographic Information Science 32.3 (2005): 139-160.

The paper also does not discuss uncertainty that comes from sparsely sampled trajectory data (extracted from social media) [8] [Krueger et al., 2016], which is similar to uncertainty caused by imprecise location estimation. 时空采样的不确定性

(not referenced) Krueger, Robert, Guodao Sun, Fabian Beck, Ronghua Liang, and Thomas Ertl. "TravelDiff: Visual comparison analytics for massive movement patterns derived from twitter." In Pacific Visualization Symposium (PacificVis), 2016 IEEE, pp. 176-183. IEEE, 2016.

3. It is important that the paper discusses the preprocessing of trajectories and the enrichment with semantic information in detail. While I can follow the general approach already, some details are not yet clear: 向量部分的解释应该更加清楚, 这里是评委不太了解为什么轨迹向量可以比较。  
\* In the vector representation, S\_T is described sufficiently, but there is no precise definition of L\_t and N\_sp.  
\* In the first level of clustering, I do not understand how vectors can be compared: As I understand the definitions, S\_T can have different length, so different vectors R\_T are not part of the same vector space.  
\* In the second level of clustering, it is unclear what "same semantics" means. Since, the semantics are assigned in a 'fuzzy' way, an equivalence relation is not obviously given.  
\* In the third level of clustering, the condition that the trajectory "commonly pass at least one region" does not lead to a partition of the trajectories into non-overlapping sets because a trajectory might share different regions with different other trajectories.
4. Hence, different types of POIs should be treated differently. I do not see this reflected in the approach, nor do I find a discussion how time information is treated exactly. POI 应该可以被区分对待, 停留和路过的区别应该更加合理化
5. However, the result is maybe to condensed already. For instance, in Figure 1, I do not see much difference between the listed trajectories, except the small number in front or behind the middle node. Is it possible to expand this number and see the full trajectory? Also, each trajectory consumes considerable space, but hundreds of trajectories might be included in a cluster. The cluster view is scrollable as the video shows, but does not have a scrollbar or indicates how many trajectories are contained in the cluster. How are the trajectories sorted in the cluster? 数量没有暗示, 列表中的元素看上去都一样, 可视化没有很好的结果。我打算这里增加一些可视化的设计。首先是用矩阵的图形来表示, 每一行是一类轨迹, 横轴时间, 关键点用红色, 其他用绿色。最后用条形图表示数量。看上去会清楚很多。





6. I do not see the need for a new and complicated hierarchy representation called HoneyComb. The nesting of hierarchy elements obfuscates the inner nodes – it is hard to see the structure of the hierarchy in the result. The overall picture of the hierarchy does not provide much helpful information; like the video shows, considerable zooming and investigating details that appear on demand is required to find meaningful information. The map-like appearance might mislead users to seek or expect a correlation of the position in the hierarchy map with the actual geographic locations. Using a plain icicle plot like done in the paper for illustration purposes might be a better solution. By the way, there is already a visualization named “Honeycomb” :  
 Van Ham, Frank, Hans-Jörg Schulz, and Joan M. Dimicco. “Honeycomb: Visual analysis of large scale social networks.” In IFIP Conference on Human-Computer Interaction, pp. 429-442. Springer Berlin Heidelberg, 2009.  
 首先要调整蜂窝图的视觉效果，强调蜂窝图的层次结构关系。然后通过比较直观的语义编码来修改这一部分
7. The paper does not describe how the trajectories and heatmaps are plotted. Given that the reconstructed trajectories are uncertain, it would be helpful to see the varying level of uncertainty in the plotted lines. One solution could be to use a splatting approach and vary the width of the kernel depending on the current level of uncertainty. I am not sure about the usefulness of the heatmaps. Do they somehow aggregate the stopover times at a certain location? How to handle uncertainty of this location?  
 地图上的轨迹和热力图，没有说明是如何展示的，并且热力图看上去没有什么作用，如果是不确定性的轨迹，要怎么展示不确定性。是否需要绑定轨迹展示
8. Further, I find it confusing that, when a query is issued, the results in the map do not show a set of trajectories already, but first just a set of POIs. 查询后的结果展示，不明显，查询流程需要重新思考设计
9. I am missing a clearer definition of the grammar in the paper. For instance, it stays unclear to what extent the predicates can be mixed and nested. Maybe, a form-based interface to specify the query would be easier to use, for instance, where I can just select time spans from a calendar, instead of having to type in the precise date. 查询输入的语法，好像需要学习，查询所选择的时间看上去很繁琐。
10. The case studies do not provide sufficient evidence for the usefulness of the approach. Some of the findings do not go beyond the trivial, for instance, “most people come to the train station for traveling “ or “that there are many facilities on education around the campus”. Even worse, many findings do not require to study any trajectory data, but are just based on locations of POIs. None of the examples convinced that the approach is effective to detect noteworthy patterns. 现有的例子，有的太模糊，而且有的甚至没有用到轨迹来分析任务。
11. I also cannot follow some of the conclusions drawn, for instance: How can I see that “Riding the train is the major way for citizens to head to other cities” (how to compare that with other means of

transport)? Or why does is the train station not satisfying? 例子的介绍不够清楚, 缺少一些分析出来的证据, 这点应该是作图的问题

12. Usually city districts and different kinds of community landuse (e.g. industry, businesses, and housing) also follow terrain restrictions and are of different shape and sizes. This is not taken into account in this work. 需不需要把多边形的大小, 形状等考虑进来, 我认为暂时不需要考虑, 还是需要在文章中说清楚, 我们的多边形是符合基站实际情况的, 介绍一下分布。
13. The size of each area (in the Voronoi partition) and the amount of POIs in each area is critical. How do you determine the granularity of the size of the areas? (R4) POI 控制范围, 和多边形一致
14. Unfortunately, there is no validation. Is may be difficult to get ground truth data (e.g. paths of people with known activities) to validate if the model also suggest the right activities. However, at least a discussion about this issue and pointing to some works that have carried our evaluation should be included! For example,  
  
S. Reumers, F. Liu, D. Janssens, M. Cools, and G. Wets. Semantic annotation of gps traces: Activity type inference. In Proceedings of the 92nd Annual Meeting of the Transportation Research Board (DVD-ROM). Transportation Research Board of the National Academies, 2013.  
  
没有 ground truth. 这部分考虑过, 可以寻找更加合理的案例, 与前人的案例进行对比, 同时经过更多的可能性度量方法来衡量可能性。
15. I think it could be difficult to understand the hierarchy, if every level is based on different decisions. The authors explain the three hierarchy levels, but it is not said why they were chosen in this order and for what purpose this is useful. 为什么要分三层. 案例用处也没有体现
16. As the items are layouted in a spiral-manner it could be that some items are next to each other which do not really have much in common. ?关于布局, 增加案例说明作用, 同时说明好处, 评估优缺点, 图片说明要更详细。
17. Another remaining question is, if the approach also finds some longer activity sequences. 希望有更加复杂的查询语句。
18. The authors should give some more background on why the studies are carried out. Is there somebody interested in the data? Where does the movement data and the POI data exactly come from? How was it derived? What are the main questions to be answered? 关于 case 需要加一些背景介绍, 动机和解决了的问题。
19. Here it remains unclear to what algorithms the approach was exactly compared to and on which machine the results were achieved. I recommend to either delete the runtime discussion completely or to add much more details. 关于性能比较, 是否需要, 还需要增加更多细节方法

I did not fully understand the encoding for the honeycomb chart. In addition to stating the way the chart was created (in section 6.2), an explanation of what the chart represents from a user (the analyst) point of view is needed. What is the encoding for the color? what does the structure of each cluster represent? how can the user "zoom in" to see a single honey comb? This is made a bit clearer in the attached movie, but needs to be explained also in the paper. 蜂窝图的交互性操作可以讲得更细节一些。

-- How did you come up with the user tasks for the querying condition? This should be based on some analysis of possible user tasks of spatial data.

User task 的动机, 目的需要加以说明

-- How does this method generalize to other types of data? What other types of data (other than cell phone GPS data) might benefit from this approach?

可拓展性加在未来工作之前可以